Hierarchical animal movement models for population-level inference†

Mevin B. Hootena*, Frances E. Budermanb, Brian M. Brostb, Ephraim M. Hanksc and Jacob S. Ivanda

New methods for modeling animal movement based on telemetry data are developed regularly. With advances in telemetry capabilities, animal movement models are becoming increasingly sophisticated. Despite a need for population-level inference, animal movement models are still predominantly developed for individual-level inference. Most efforts to upscale the inference to the population level are either post hoc or complicated enough that only the developer can implement the model. Hierarchical Bayesian models provide an ideal platform for the development of population-level animal movement models but can be challenging to fit due to computational limitations or extensive tuning required. We propose a two-stage procedure for fitting hierarchical animal movement models to telemetry data. The two-stage approach is statistically rigorous and allows one to fit individual-level movement models separately, then resample them using a secondary MCMC algorithm. The primary advantages of the two-stage approach are that the first stage is easily parallelizable and the second stage is completely unsupervised, allowing for an automated fitting procedure in many cases. We demonstrate the two-stage procedure with two applications of animal movement models. The first application involves a spatial point process approach to modeling telemetry data, and the second involves a more complicated continuous-time discrete-space animal movement model. We fit these models to simulated data and real telemetry data arising from a population of monitored Canada lynx in Colorado, USA. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: hierarchical model; resource selection model; spatial statistics; telemetry data; trajectories

1. INTRODUCTION

The field of movement ecology is booming, in large part, because of the increased availability of telemetry data sources (Cagnacci et al., 2010). Contemporary telemetry data are acquired via satellite communication devices affixed to individual animals. These devices often collect many types of data, but most studies are focused on the position data, primarily to learn about environmental influences on individual-level movement. Many new statistical models for animal trajectories have been proposed in recent years, and they vary in form depending on the motivation for the project and type of inference desired (Hooten et al., in press). For example, most individual-based statistical models for telemetry data fall into one of three classes: point process models, discrete-time models, or continuous-time models, with each being appropriate in certain settings (McClintock et al., 2014).

Statistical inference arising from fitting animal movement models to telemetry data is sometimes focused on the individual level. For example, a movement ecologist might ask how a specific individual animal responded to environmental cues while migrating between summer and winter home ranges (e.g., Hooten et al., 2010a). However, many animal movement studies are concerned with population-level inference. That is, for several individuals, is there evidence of consistent behavioral responses to environmental variables? To obtain population-level inference, the well-accepted approach is to use a hierarchical model with random effects for individuals that are pooled at the population level. For example, consider the Bayesian hierarchical model

\[ y_j \sim [y_j | \beta_j, \theta_j] \]

†This material is associated with the President’s Invited Lecture from the 26th Annual Conference of the International Environmetrics Society (TIES), presented 18 July 2016.

Correspondence to: M. B. Hooten, U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit; Departments of Fish, Wildlife, & Conservation Biology and Statistics, Colorado State University, Fort Collins, CO 80523 U.S.A. E-mail: mevin.hooten@colostate.edu

U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit; Departments of Fish, Wildlife, & Conservation Biology and Statistics, Colorado State University, Fort Collins, CO 80523 U.S.A.

Department of Fish, Wildlife, and Conservation Biology, Colorado State University

Department of Statistics, Pennsylvania State University

Colorado Parks and Wildlife
The complexity of modern animal movement models makes implementation challenging. Furthermore, increases in the quantity of data resulting from newer telemetry devices has outpaced computational methods for fitting animal movement models. Animal ecologists may wish to extend individual-level models to provide statistically rigorous population-level inference, but, in many cases, the algorithms required to fit such models become prohibitively challenging to program or are too slow in settings with large data sets and/or many individuals. For example, Hanks et al. (2011) performed a post hoc meta-analysis to obtain population-level inference for northern fur seals (Callorhinus ursinus) because the implementation of a full hierarchical movement model was not computationally feasible. Furthermore, in the Bayesian setting, Markov Chain Monte Carlo (MCMC) algorithms for most animal movement models require tuning from the user due to lack of conjugacy. In cases where data sets from tens or hundreds of individuals are available, it may not be feasible to tune individual-level Metropolis-Hastings updates for all parameters.

We present a statistically rigorous two-stage procedure for economizing hierarchical animal movement models to provide exact population-level inference using a sequence of algorithms that are fast, stable, and require little or no tuning by the user. Our approach is simple. First, we fit individual-level models (1) independently using a preferred stochastic sampling algorithm. Independent model fits in the first stage allow for parallel processing, leading to an improvement in computational efficiency that scales with the number of processors. Second, we obtain exact population-level inference using a secondary MCMC algorithm that requires no tuning. The secondary algorithm is based on a little-known technique for Bayesian meta-analysis proposed by Lunn et al. (2013). We found that our two-stage procedure provides substantial computational improvements in both speed and ease of use in cases with large data sets and/or complicated data models.

In what follows, we present a general two-stage procedure for fitting a broad class of hierarchical animal movement models. We then demonstrate the approach for a basic point process model for telemetry data (i.e., resource selection function model) and verify it using simulation. In our second application, we show how the approach can be applied to a continuous-time discrete-space (CTDS) animal movement model using telemetry data with complicated error structure. We apply the CTDS model to satellite telemetry data from a population of Canada lynx (Lynx canadensis) in Colorado, USA. Finally, we close with a summary and discussion of the approach and future directions.

2. TWO-STAGE PROCEDURE

Many animal movement models have been constructed solely for individual-level inference (e.g., Jonsen et al., 2005; Johnson et al., 2008b; Hooten et al., 2010a; Brost et al., 2015; Buderman et al., 2016). However, the desired scientific inference is usually at the population level to assess if the population, as a whole, is responding to certain environmental cues. Hierarchical statistical models provide a natural framework for obtaining upscaled population-level inference (Gelman and Hill, 2006; Hobbs and Hooten, 2015). As the complexity of animal movement models increases, hierarchical models that include nonlinear components become challenging to implement due to computational limitations and user supervision requirements. It is often much simpler to fit individual-level models to data, as long as individuals are assumed independent. Following Lunn et al. (2013), we propose a simple two-stage procedure for obtaining population-level inference under the full hierarchical model. The two-stage procedure only requires independent individual-level model fits and an unsupervised resampling algorithm to obtain population-level inference without any user tuning.

The first stage in the procedure involves fitting a data model like (1) independently for each individual \( j = 1, \ldots, J \). In addition to the prior for auxiliary data-level parameters \( \theta_j \) from (5), we also specify a prior for the individual-level parameters \( \beta_j \) as \( \beta_j \sim \beta \) (where the priors for \( \theta_j \) and \( \beta_j \) can differ by individual). The priors for \( \beta_j \) are only used in the first stage of the two-stage procedure and do not affect the final inference. The posterior distribution for individual \( j \) is

\[
[\theta_j, \beta_j | y_j] = \frac{[y_j | \beta_j, \theta_j] \beta_j}{\int [y_j | \beta_j, \theta_j] \beta_j d\beta_j d\theta_j}.
\]

In principle, any stochastic sampling algorithm can be used to obtain samples from the posterior distribution in (6), but those relying on MCMC are most commonly applied in the animal movement literature. However, because we treat the models in (6) for all \( J \) individuals independently in the first stage, they can be fit in parallel using readily available software (e.g., the ‘parallel’ R package; R Core Team, 2016). Additionally, if we choose a sampling algorithm for fitting the models in (6) that is unsupervised (i.e., requiring no supervised tuning), then the entire two-stage procedure can be automated. An unsupervised fitting procedure will be used much more often by ecologists in...
situations where data exist for a large number of individuals. Thus, automatic MCMC algorithms like BUGS (Lunn et al., 2009), JAGS (Plummer, 2003), or STAN (Carpenter et al., 2016) can be used to fit the individual-level models in (6), or alternatively, importance sampling or particle filtering (e.g., LibBi; Murray, 2013) methods can also be employed. Finally, the choice of priors \( \{ \beta_j \} \) can also lead to fully automatic and parallelizable first-stage algorithms. For example, if the data model (1) is Poisson (i.e., \( y_j \sim \text{Pois}(\exp(X_j \beta_j)) \)), where \( X_j \) is a design matrix of covariates for the \( j \)th individual, then \( \beta_j \) is empty because the Poisson does not have a separate dispersion parameter.

A multivariate log-gamma prior distribution (Crooks, 2010; Bradley et al., 2015) for \( \beta_j \) facilitates the use of a Monte Carlo sampler to obtain posterior samples from (6). For non-conjugate priors, adaptively tuned MCMC algorithms (e.g., Givens and Hoeting, 2012) are straightforward to implement and provide a way to obtain unsupervised stage-one samples for \( \beta_j \).

The second stage in the two-stage procedure involves an MCMC algorithm resembling that used to fit the full hierarchical model but with a critical simplification. To fit the full hierarchical model in (1)–(5), we sequentially sample from the full-conditional distributions \( \{ \beta_j \} \) for \( j = 1, \ldots, J \), \( \{ \mu_\beta \} \), and \( \{ \Sigma_\beta^{-1} \} \), using an MCMC algorithm. In our second stage algorithm, we use the MCMC algorithm for the full hierarchical model as a template but modify the updates for \( \beta_j \). Updates for the individual-level auxiliary parameters, \( \theta_j \), are automatically coupled with those from \( \beta_j \) but are only necessary if we desire inference for \( \theta_j \). In fact, if \( \theta_j \) are considered nuisance parameters, it is not necessary to store samples for them in our two-stage procedure.

The full-conditional distributions for population-level parameters \( \mu_\beta \) and \( \Sigma_\beta^{-1} \) in the second stage model remain the same as in the MCMC algorithm to fit the full hierarchical model in (1)–(5):

\[
[\mu_\beta | \cdot] \propto \left( \prod_{j=1}^{J} [\beta_j | \mu_\beta, \Sigma_\beta] \right) [\mu_\beta].
\]

(7)

\[
[\Sigma_\beta^{-1} | \cdot] \propto \left( \prod_{j=1}^{J} [\beta_j | \mu_\beta, \Sigma_\beta] \right) [\Sigma_\beta^{-1}].
\]

(8)

If the model for \( \beta_j \) and prior for \( \mu_\beta \) are multivariate Gaussian and the prior for \( \Sigma_\beta^{-1} \) is Wishart, then the full-conditional distributions in (7) and (8) are multivariate Gaussian and Wishart, respectively. These specific distributions are commonly used in many animal movement models for population-level parameters and permit conjugate Gibbs updates in our second-stage algorithm.

The joint full-conditional distribution for the data-level auxiliary parameters, \( \theta_j \), and individual-level parameters, \( \beta_j \), is

\[
[\theta_j, \beta_j | \cdot] \propto [y_j | \beta_j, \theta_j] [\beta_j | \mu_\beta, \Sigma_\beta] [\theta_j | \mu_\beta, \Sigma_\beta]
\]

(9)

which, depending on the form of data model \( [y_j | \beta_j, \theta_j] \), would normally require a Metropolis-Hastings update. In this case, the Metropolis-Hastings ratio for the joint update of \( \theta_j \) and \( \beta_j \) is

\[
\frac{r_j}{\frac{[y_j | \beta_j^*, \theta_j^*] [\beta_j^* | \mu_\beta^*, \Sigma_\beta^*] [\theta_j^* | \mu_\beta^*, \Sigma_\beta^*]}{[y_j | \beta_j^{k-1}, \theta_j^{k-1}] [\beta_j^{k-1} | \mu_\beta^{k-1}, \Sigma_\beta^{k-1}] [\theta_j^{k-1} | \mu_\beta^{k-1}, \Sigma_\beta^{k-1}]}}
\]

(10)

where, the ‘*’ superscript represents the proposal for \( \beta_j \) and the ‘k’ and ‘k − 1’ superscripts correspond to the MCMC sample for the \( k \) or \( k – 1 \) iteration of the MCMC algorithm (for \( k = 2, \ldots, K \)). Typically, the proposal distribution, \( [\theta_j^*, \beta_j^* | \theta_j^{k-1}, \beta_j^{k-1}] \), is chosen to be a multivariate Gaussian random walk such that \( (\theta_j^*, \beta_j^*)' \sim N((\theta_j^{k-1}, \beta_j^{k-1})', \Sigma_j) \) that requires tuning for each individual \( j \) by adjusting \( \Sigma_j \) using trial and error or an adaptive MCMC approach (e.g., Roberts and Rosenthal, 2009).

However, if we use the posterior samples for \( \theta_j \) and \( \beta_j \) from the first stage (6) as the proposal in the second stage update for \( \beta_j \), then the proposal distribution is

\[
[\theta_j^*, \beta_j^* | \theta_j^{k-1}, \beta_j^{k-1}] = \frac{[y_j | \beta_j^*, \theta_j^*] [\beta_j^* | \mu_\beta^*, \Sigma_\beta^*] [\theta_j^* | \mu_\beta^*, \Sigma_\beta^*]}{[y_j | \beta_j | \theta_j] [\beta_j | \mu_\beta, \Sigma_\beta] [\theta_j | \mu_\beta, \Sigma_\beta]}
\]

(11)

which does not depend on the previous \( \beta_j^{k-1} \) and \( \theta_j^{k-1} \). The Metropolis-Hastings ratio from (10) simplifies to

\[
r_j = \frac{[\beta_j^* | \mu_\beta^*, \Sigma_\beta^*] [\beta_j^{k-1}]}{[\beta_j | \mu_\beta, \Sigma_\beta] [\beta_j^{k-1}]}
\]

(12)

while the updates for \( \mu_\beta \) and \( \Sigma_\beta^{-1} \) remain unchanged. Thus, we keep the samples for \( \theta_j^* \) and \( \beta_j^* \), from the first stage, with probability \( \min(r_j, 1) \). However, we only need to explicitly save samples for the auxiliary individual-level parameters \( (\theta_j) \) in the first or second stages if we desire inference on them because \( r_j \), from (12), does not depend on \( \theta_j \). Furthermore, Lunn et al. (2013) note that, when the stage one priors for \( \beta_j \) are diffuse, the ratio simplifies further to \( r_j = \frac{[\beta_j^* | \mu_\beta^*, \Sigma_\beta^*]}{[\beta_j | \mu_\beta, \Sigma_\beta]} \), a mere quotient involving the
individual-level process distributions. However, we retain the form in (12) so that we can use prior information when available. Because there is no Markov dependence in the proposal for $\beta_j$, we select $\beta^*_j$ and $\theta^*_j$, if desired, uniformly at random from the output resulting from the first-stage model fits. More importantly, the Metropolis-Hastings ratios ($r_j$, for $j = 1, \ldots, J$) in (12) do not contain a tuning parameter, resulting in unsupervised updates. Paired with the Gibbs updates for $\mu_\beta$ and $\Sigma_\beta^{-1}$, the second-stage algorithm is fully automatic, and samples from the full-conditional for $\beta_j$ can be obtained in parallel (within the broader second-stage MCMC algorithm) creating the potential for additional computational efficiency. Critically, the Metropolis-Hastings ratio, $r_j$ in (12), is not a function of the data. Therefore, compicated data models do not need to be reconsidered in the second-stage algorithm. The utility of the simple two-stage procedure is that it is intuitive, facilitates parallelization, and can result in algorithms that are fully automatic.

In what follows, we provide two example applications where the two-stage procedure for obtaining population-level animal movement inference is valuable. The first application involves a spatial point process modeling approach for telemetry data commonly referred to as “resource selection function” (RSF) analysis (e.g., Manly et al., 2007). The second application involves a continuous-time discrete-space animal movement model proposed by Hooten et al. (2010a) and Hanks et al. (2015a).

3. APPLICATIONS

3.1. Hierarchical point process models

Perhaps the most common model fit to temporally independent telemetry data is the RSF model. The RSF model is a heterogeneous point process model that conditions on the number of telemetry observations. Assuming there is no measurement error associated with the telemetry data $s_{ij}$ (typically a $2 \times 1$ vector) for observations $i = 1, \ldots, n_j$ and individuals $j = 1, \ldots, J$, the data model takes the form of a weighted distribution (Patil and Rao, 1977) such that $s_{ij} \sim [s_{ij} | \beta_j]$ and

$$[s_{ij} | \beta_j] = \frac{g(x(s_{ij}), \beta_j) f(s_{ij})}{\int g(x(s), \beta_j) f(s) ds}$$

where $g(x(s), \beta_j)$ is the “selection” function and $f(s)$ is the “availability” function. Thus, the animal movement interpretation of (13) is that inference for $\beta_j$ provides insight about how individual $j$ selects resources (i.e., covariates, $x$) from those available to it. The selection function is often chosen to be exponential (i.e., $g(x(s_{ij}), \beta_j) = \exp(x(s_{ij}) \beta_j)$) and the availability function is typically assumed to be uniform on the support of the point process (i.e., $f(s_{ij}) = \text{unif}(S)$ for $s_{ij} \in S \subset \mathbb{R}^d$).

Warton and Shepherd (2010) and Aarts et al. (2012) showed that the RSF model in (13) can be fit using a variety of approaches, including a Poisson likelihood. The Poisson likelihood can be considered by first preprocessing the data such that $y_j = (y_{1,j}, \ldots, y_{m,j})$ represents counts of telemetry locations in grid cells corresponding to a discretization of the support $S$. As the grid cell size decreases with respect to the resolution of the covariates $x$, a Poisson data model coincides with the point process model. Thus, the corresponding hierarchical model

$$y_j \sim \text{Pois}(\exp(x_j \beta_j))$$
$$\beta_j \sim N(\mu_\beta, \Sigma_\beta)$$
$$\mu_\beta \sim N(\mu_0, \Sigma_0)$$
$$\Sigma_\beta^{-1} \sim \text{Wish}(Sv)^{-1}, v)$$

assumes the same form as (1)–(5) and allows for population-level resource selection inference on $\mu_\beta$. To fit the full hierarchical model directly using MCMC, we sample from the full-conditional distributions for $\beta_j$, $\mu_\beta$, and $\Sigma_\beta^{-1}$, sequentially. Standard Metropolis-Hastings updates for $\beta_j$ require tuning, but the model can be fit using a single MCMC algorithm for moderately sized data sets. Alternatively, the weighted least-squares proposal approach of Gamerman (1997) could be used to acquire samples for $\beta_j$ from the posterior distribution. However, to adequately approximate the point process model, the grid cells often need to be quite small, resulting in a fine-scale discretization of the support $S$ and increasing the computational burden.

The two-stage procedure we described in the previous section can easily be employed to fit the hierarchical model in (14)–(17). For the first stage, we can use an MCMC or Hamiltonian Monte Carlo algorithm (via BUGS, JAGS, or STAN; Lunn et al., 2009; Plummer, 2003; Carpenter et al., 2016) to fit the individual level models in parallel. For our spatial point process setting, the individual-level models are

$$y_j \sim \text{Pois}(\exp(x_j \beta_j))$$
$$\beta_j \sim N(\mu_0, \Sigma_0)$$

for $j = 1, \ldots, J$, independently. Note that the individual-level parameter model in (19) is an exchangeable prior for all $j = 1, \ldots, J$. Also, if the individual data sets $y_j$ and $X_j$ are so large that they are difficult to store in memory simultaneously for all $J$ individuals, the first-stage model fitting can be fully distributed among separate machines or performed in sequence. This highlights another primary advantage of the two-stage procedure.
The second-stage algorithm for obtaining population-level inference is an MCMC algorithm with Gibbs updates for $\mu_\beta$ and $\Sigma_\beta^{-1}$ as described in the previous section and updates for $\beta_j$ using Metropolis-Hastings based on the acceptance ratio in (12), which becomes

$$r_j = \frac{N \left( \beta_j^* \mid \mu_\beta, \Sigma_\beta \right) N \left( \beta_j^{k-1} \mid \mu_0, \Sigma_0 \right)}{N \left( \beta_j^{k-1} \mid \mu_\beta, \Sigma_\beta \right) N \left( \beta_j^* \mid \mu_0, \Sigma_0 \right)}.$$  \hfill (20)

Within the second stage MCMC algorithm, the updates for $\beta_j$ can also be parallelized because they are independent, although this model is simple enough that parallelization is not necessary in the second stage algorithm. Thus, the data, $y_j$ for $j = 1, \ldots, J$, which could include counts for 10s or 100s of thousands of grid cells and 100s of individuals, do not appear in the second stage algorithm. The absence of $y_j$ leads to a more computationally efficient second stage algorithm than the original algorithm to fit the full hierarchical model directly.

We simulated point process data from 20 individuals (Figure 1), resulting in approximately 30 simulated telemetry fixes per individual, and fit the hierarchical RSF model using: 1.) a single MCMC algorithm, and 2.) our two-stage procedure. We compared the population-level results from the fits resulting from each procedure.

For the first-stage algorithm in our two-stage procedure, we fit the individual-level models independently using an adaptive MCMC algorithm in parallel using R (R Core Team, 2016) and assumed $N(0, 100^{-1})$ priors for $\beta_j$, a $N(0, 100^{-1})$ prior for $\mu_\beta$, and a Wish($3, 1^{-1}$) prior for $\Sigma_\beta^{-1}$. Our first-stage algorithm uses a multivariate Gaussian proposal for $\beta_j$ and adapts the tuning using a single variance parameter, resulting in an unsupervised algorithm for the individual-level model fits. We could have also used BUGS or JAGS to fit the first-stage models, but our adaptive MCMC algorithm required less computing time.

The single MCMC algorithm to fit the full hierarchical model required 2.62 min to obtain 20,000 MCMC samples in R, whereas the first-stage algorithm required 0.57 min to obtain the same number of samples using an adaptive MCMC algorithm in parallel for the 20 individuals. The second-stage algorithm required only 1.49 min in R, which implies that the total computer time to fit the model using the two-stage procedure was 2.06 min (0.56 min less than the single MCMC algorithm). Also, the two-stage procedure requires no tuning and results in much larger effective MCMC sample sizes for parameters. The effective MCMC sample sizes for $\mu_\beta$ and $\beta_j$ were 8560 and 1398 (averaged across individuals) for the single MCMC algorithm, but were 17,590 and 15,184 for the two-stage algorithm (out of 20,000 total samples). Thus, to obtain the same effective MCMC sample size using MCMC for all parameters, we would need an order of magnitude more samples from the single MCMC algorithm.

Figure 2 illustrates the similarities in inference for the slope parameters $\mu_{\beta_3}$ and $\beta_{j1}$ for $j = 1, \ldots, 20$ when fitting the hierarchical RSF model using a single MCMC algorithm (black) versus the two-stage procedure (gray). Notice that the single MCMC algorithm and the two-stage procedure provide very similar inference. In terms of inference, there exists some variability among individuals, but the population-level inference (Figure 2, top) suggests a consistent overall positive population response to the covariate.

### 3.2. Hierarchical continuous-time discrete-space models

The previous application, involving spatial point process models, involves a commonly used model specification and desired type of inference in ecological research, but more contemporary methods have been developed to explicitly model the dynamics of animal movement based on temporally dependent telemetry data with observations close in time. Among these methods are discrete-time and continuous-time approaches to modeling the individual animal trajectories (McClintock et al., 2014). We focus on the continuous-time class of models in what follows.
Continuous-time statistical models for animal movement processes have existed for decades (e.g., Dunn and Gipson, 1977; Blackwell, 1997) and are usually based on Brownian motion (i.e., Wiener processes). Up until the late 1990s, most Brownian motion models for trajectories utilized an Ornstein–Uhlenbeck process (i.e., a Wiener process with attraction to a central position). Johnson et al. (2008b) also proposed an Ornstein–Uhlenbeck model but for the velocity (i.e., temporally differentiated position) rather than the position process. Hooten and Johnson (2016a) generalized the continuous-time velocity models of Johnson et al. (2008b) in the context of Gaussian processes with covariance structure induced by temporal basis functions. Buderman et al. (2016) used a simplified basis function parameterization to model Canada lynx (Lynx canadensis) movement while accounting for measurement error in the telemetry data. Buderman et al. (2016) refer to their model as a “functional movement model” and use it to provide inference for the true underlying continuous position process (i.e., $\mu(t)$, for time $t$) of an individual.
The approach developed by Buderman et al. (2016) assumes that the telemetry data \(s_{ij}\) are observed with error. In fact, for the Canada lynx in our study, the bivariate measurement error follows an unusual X-shaped pattern because the telemetry data are collected by Service Argos (Costa et al., 2010) that relies on polar orbiting satellites. Thus, Brost et al. (2015) and Buderman et al. (2016) developed a measurement error model based on a mixture distribution to account for the X-shaped Argos pattern (see Appendix A for details). Properly accounting for measurement error adds another level to the hierarchical model in (1)–(5) such that

\[
\begin{align*}
  s_{ij} &\sim \begin{cases} 
  [s_{ij}|\mu_j(t_l), \phi_j] 
  
  \end{cases} 
  \tag{21} 
  
  y_{ij} &\sim [y_{ij}|\beta_j, \theta_j] 
  \tag{22} 
  
  \beta_j &\sim [\beta_j|\mu_\beta, \Sigma_\beta] 
  \tag{23} 
  
  \mu_\beta &\sim [\mu_\beta] 
  \tag{24} 
  
  \Sigma_\beta^{-1} &\sim \begin{cases} 
  \Sigma_\beta^{-1} 
  \end{cases} 
  \tag{25} 
  
  \theta_j &\sim [\theta_j] 
  \tag{26} 
  
  \phi_j &\sim [\phi_j] 
  \tag{27}
\end{align*}
\]

for \(j = 1, \ldots, J\) individuals, and where \(y_{ij}\) is an \(m_j \times 1\) vector that represents a latent process that is linked to the true continuous position process \([\mu_j(t), \forall t]\) by a deterministic functional \(h\) such that \(y_{ij} = h([\mu_j(t), \forall t])\) and \(\phi_j\) are measurement error covariance parameters.

Hooten et al. (2010a) developed an individual-level animal movement model based on (21) and (22) where the latent variables \(y_{ij}\) represent a sequential multinomial process indicating transitions among grid cells on a discretization of the spatial support \(S\). The latent process model in (22) relies on a continuous-time discrete-space (CTDS) representation of the position process. However, because the functional \(h(t)\), which links the position process with the data, is non-invertible in their model, Hooten et al. (2010a) proposed a Bayesian multiple imputation procedure to account for uncertainty in the true position process when making inference on \(\beta_j\). The multiple imputation procedure used by Hooten et al. (2010a) differs from the two-stage procedure described herein because it does not allow for feedback from the individual-level parameters \(\beta_j\) to the position process \([\mu_j(t), \forall t]\) or measurement error parameters \(\phi_j\). Hooten et al. (2010a) used an imputation model to interpolate the position process and then integrated over the uncertainty in the position process while fitting (22) to provide posterior inference for the individual-level parameters \(\beta_j\).

Hanks et al. (2015a) showed that the multinomial process of Hooten et al. (2010a) could be reparameterized such that \([y_{ij}|\beta_j, \theta_j]\) can be modeled using Poisson regression. Specifically, let \(\tau_{ij}\) represent the amount of time individual \(i\) remains in a grid cell for the \(c\)th “stay/move” pair associated with the discretization of the individual’s path through a landscape (for \(c = 1, \ldots, n_j\)). Then let \(y_{cij} \sim \text{Pois}(\tau_{ij} \exp(\mathbf{x}_{cij}^T \beta_j))\) where the index \(l = 1, 2, 4, 5, 6, 7\) is not necessary because it corresponds to the middle cell that is captured by \(\tau_{ij}\) denotes moves to neighboring grid cells in each cardinal direction (i.e., north, east, south, west). That is, if individual \(i\) moved north for “stay/move” pair \(c\), then the data point \(y_{c1j} = 1\) and \(y_{c2j} = y_{c3j} = y_{c4j} = 0\) (see Appendix B for details). The Poisson reparametrization dramatically improves computational efficiency at the individual level because the total number of observations used in the model \((4m_j)\) is a function of the grid cell size rather than the position process discretization as used in Hooten et al. (2010a). Thus, Hanks et al. (2015a) were able to fit the CTDS model to large telemetry data sets in a fraction of the time required by the multinomial method developed by Hooten et al. (2010a). However, neither Hooten et al. (2010a) nor Hanks et al. (2015a) attempted to fit a hierarchical model like that in (21)–(27) to obtain population level inference for \(\mu_\beta\).

In our application involving population-level inference for Canada lynx, we use the model developed by Buderman et al. (2016) to obtain the imputation distribution for the true individual-level position process \([\mu_j(t), \forall t]\) and hence \(\tilde{y}_{cij}\) for all \(c, l, j\), while accounting for the complicated nature of Argos telemetry error (see Appendix A for details). In what follows, we combine all \(\tilde{y}_{cij}\) into a single vector representing the latent process \(\tilde{y}_j\) and use \(\tilde{y}_j\) as data in a two-stage implementation of the hierarchical model in (21)–(27).

To fit the hierarchical model using the two-stage procedure described in Section 2, we apply the same two stages of algorithms as in the previous application. For the first stage, we use the data model in (22) and specify multivariate Gaussian priors for the individual-level parameters \(\beta_j \sim N(\mu_\beta, \Sigma_\beta)\). We use an adaptively tuned MCMC algorithm to obtain samples from the posterior distributions

\[
[\beta_j|s_{ij}, \forall i, j] = \int \left[ \beta_j|\tilde{y}_j \right] [\tilde{y}_j|s_{ij}, \forall i, j] \, d\tilde{y}_j 
\]

for \(j = 1, \ldots, J\), and where \([\tilde{y}_j|s_{ij}, \forall i, j]\) represents the imputation distribution for the latent Poisson process. To perform the integration in (28), we sample \(\tilde{y}_j^k \sim [\tilde{y}_j|s_{ij}, \forall i, j]\) on the \(k\)th MCMC iteration and then let the Metropolis-Hastings update \(\beta_j^k\) depend on \(\tilde{y}_j^k\) as described in Hooten et al. (2010a) and Hanks et al. (2015a). As in the first application, we can fit the \(J\) models for all individuals in parallel, dramatically reducing the required computational time.
Figure 3. (a) Colorado, USA, with major highways and the city of Denver shown. The telemetry data spanning a year of time for 18 individual Canada lynx are shown as points. A shaded relief map is shown as the background image to illustrate the topography of the area. (b) and (c) close-up views of the predicted paths for two individual Canada lynx. For clarity, only the posterior mean path is shown.

For the second stage of the two-stage procedure, we use the posterior samples for \( \{ \beta_j, \forall j \} \), from the first stage, as proposals in the MCMC algorithm to fit the hierarchical model in (22)–(25). In doing so, we update \( \{ \beta_j, \forall j \}, \mu_\beta, \) and \( \Sigma^{-1}_\beta \) sequentially in a completely unsupervised second-stage MCMC algorithm. Recall that the Metropolis-Hastings acceptance ratio for \( \beta_j \) is identical to that used in the previous application (20). As a result of the two-stage implementation and the adaptive tuning in the first-stage algorithm, the procedure is completely automatic after the data are preprocessed to obtain the imputation distribution, and population-level inference for \( \mu_\beta \) can easily be obtained.

Using telemetry data from \( J = 18 \) individual Canada lynx in Colorado, USA (Figure 3(a)), we applied the two-stage procedure to fit the hierarchical model in (22)–(25). We used the functional movement model of Buderman et al. (2016) to obtain the imputed path distribution (Figure 3(b,c)) for each individual and used nearly continuous imputed path realizations to create the latent Poisson data realizations \( y^k_j \) (resulting in approximately 450 discrete-space transitions per individual, \( n_j \approx 450 \)). Canada lynx are a subalpine species that tend to prefer forested ecosystems (McKelvey et al., 2000); thus, we focused on two covariates: elevation and distance to forest (Figure 4). Each covariate was included in the model as a “static” driver, rather than a gradient-based driver of movement (Hanks et al., 2015a). Static drivers can be interpreted as affecting overall motility in the CTDS model. For priors in the first stage, we used \( \beta_j \sim N(0, 100I) \) for all \( j = 1, \ldots, 18 \). We used \( \mu_\beta \sim N(0, 100I) \) and \( \Sigma^{-1}_\beta \sim \text{Wish}(3 \cdot I)^{-1}, 3 \) as priors for the population-level parameters and precision matrix. See Appendix B for additional details on the CTDS animal movement model.
Figure 4. Images of covariates with telemetry observations overlaid as black points: (a) Elevation and (b) distance to forest. Light shading corresponds to larger values

Figure 5. Posterior estimates for the population-level parameters $\mu_B$ and individual-level parameters $\beta_j$. The posterior mean is shown as a central point, and the 50% and 95% credible intervals are shown as the thick and thin black lines. (a) Results for the elevation covariate and (b) results for the distance from forest covariate.

We fit the overall hierarchical model using the two-stage procedure, and the resulting algorithms required 0.86 min for the first stage (using an adaptive MCMC algorithm in parallel) and 1.62 min for the second stage. Figure 5 shows the results of the model fit in terms of posterior means, and 50% and 95% credible intervals for the population-level parameters $\mu_B$ and individual-level parameters $\beta_j$.

While there exists substantial variability among individual Canada lynx, with some individuals exhibiting clear relationships with the covariates (e.g., individuals 2, 4, and 5), the posterior distributions for $\mu$ did not indicate a population-level effect for either covariate at the 95% level (but both did at the 50% level). For the individuals that did show evidence of an effect (i.e., 95% credible intervals not overlapping zero), the negative response to elevation indicates that overall motility decreases at higher elevations, leading to greater residence times in those regions, as opposed to lower elevations (Figure 5(a)). Similarly, for individuals with significant effects related to distance from forest we see positive influence on motility implying that those Canada lynx have higher motility (and hence lower residence time) in regions farther from forest (Figure 5(b)). Thus, the inference in our application involving Canada lynx agrees with that obtained in other studies (e.g., McKelvey et al., 2000).

4. CONCLUSION

Our findings indicate that the two-stage procedure we described herein holds tremendous value for fitting hierarchical animal movement models to telemetry data for population-level inference. We applied the two-stage procedure to two types of commonly used animal movement models of varying complexity and found that it worked well in both cases.

The spatial point process modeling approach we described in the first application is a commonly used model, but still fairly simple. Much more complicated spatio-temporal point process models have been used to model temporally correlated telemetry data (e.g., Johnson et al., 2008a; Johnson et al., 2013; Brost et al., 2015), and adapting the two-stage procedure to those models is the subject of ongoing research. For example, Brost et al. (2015) developed a model with a time-varying dynamic availability component that depended on an additional smoothness parameter. Thus, the data model developed by Brost et al. (2015) required substantially more computation time than...
the simulated example we presented in Section 3.1 and would benefit from a two-stage implementation where individual-level models could be fit independently on separate processors and then recombined using the second-stage MCMC algorithm to yield population-level inference for $\mu_G$.

In our example involving Canada lynx, the continuous-time discrete-space reparameterization developed by Hanks et al. (2015a) already provides significant improvements in computational efficiency over the motivating model developed by Hooten et al. (2010a). However, additional computational gains can be achieved using the two-stage fitting procedure to provide population-level inference.

Despite the wide range of potential applications to many types of hierarchical models, we found it surprising that the two-stage fitting procedure of Lunn et al. (2013) is not more well known. For our situations with large amounts of telemetry data and potentially complicated data models, we found the two-stage procedure works very well and is trivial to implement. We also found it very helpful to be able to use different data models, first-stage fitting algorithms, and easy parallelization. As a potential caveat, the two-stage procedure described by Lunn et al. (2013) may not be very efficient when the population induces extreme amounts of shrinkage in the individual-level parameters. Thus, in these cases, more samples would be needed in the first-stage algorithm. However, in a preliminary simulation study, we found that the two-stage procedure performs poorly only for data sets with very small amounts of data (i.e., < 20 observations for a subset of individuals).

Animal movement models have also been developed to account for more mechanistic interactions among individuals (e.g., Russell et al., 2016; Scharf et al., 2015), and while we did not address those specifically, the approach we presented may also be beneficial in those settings. Furthermore, Bayesian animal movement models have been fit using integrated nested Laplace approximation (INLA; Rue et al., 2009; Illian et al., 2012; Illian et al., 2013; Ruiz-Cárdenas et al., 2012; Jonsen, 2016), and one could use INLA to fit the hierarchical point process model in our first example. However, the two-stage MCMC approach presented herein allows for the following: inference on joint relationships among model parameters, easy parallelization in the first stage, and the ability to use Bayesian multiple imputation techniques, such as in our second example involving the CTDS movement model.

Acknowledgements

Support for this research was provided by NSF 1614392, NSF EEID 1414296, CPW T01304, and NOAA AKC 188000. The authors thank The International Environmetrics Society and the editors of Environmetrics for their support and assistance with this work. The authors also thank Walt Piegors, Mindy Rice, Devin Johnson, Peter Craigile, Erin Peterson, Ron Smith, and the other organizers of the TIES 2016 annual meeting. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES


**APPENDIX A. THE IMPUTATION DISTRIBUTION**

Buderman *et al.* (2016) developed a phenomenological statistical model for estimating an individual’s underlying continuous-time path based on Argos telemetry data and a semiparametric regression using temporal basis functions. We used this model to precalculate an imputation distribution for the true path. For the $j$th individual, the FMM developed by Buderman *et al.* (2016) is
movement rates can then be linked to the environmental covariates by a log-linear link which implies that.

\[
\mu_j(t_i) = W_j(t_i)\alpha
\]

\[
\alpha \sim N(0, \Sigma_\alpha)
\]

where \(s_{ij}\) represents the \(i\)th telemetry observation, \(\mu_j(t_i)\) is the true individual position at time \(t_i\), \(\Sigma\) is an error covariance matrix on the first axis, and \(HE_\Sigma, H\) is the error covariance matrix on a rotated axis (\(H\) is a rotation matrix). The probability \(p\) allows the telemetry data to arise from a bivariate Gaussian mixture that captures the X-shaped error pattern inherent to Argos data. The matrix \(W_j(t_i)\) contains basis vectors (i.e., b-spline basis vectors) at time \(t_i\) for individual \(j\), and \(\alpha\) is a set of regression coefficients corresponding to the temporal basis functions. Buderman et al. (2016) set \(\Sigma_\alpha = \text{Diag}(\sigma^2_\alpha)\) and tuned \(\sigma^2_\alpha\) to induce regularization in the model and improve predictive ability (i.e., ridge regression).

The imputed path distribution is obtained by sampling from the posterior predictive distribution of \([\mu_j(t)|[s_{ij}, \forall i, j]]\) for a large, but finite, set of times \(t \in T\) to obtain posterior realizations \(\mu^k(t)\) for \(k = 1, \ldots, K\) MCMC iterations. Figure A1(a) shows an example set of path realizations (lines) that could result from fitting the FMM from Buderman et al. (2016) to telemetry data (points).

Figure A1(b) shows a zoomed in section of the path realizations that highlight the temporal discretization. At a finer spatial resolution, we can see that the path realizations cross through an example grid cell and its associated neighborhood (Figure A1(c)). This idea is critical for processing the path realizations for use with the CTDS model.

### APPENDIX B. CTDS MODEL

For each individual \(j\) in the original CTDS model, each segment (between points) in Figure A1(c) served as a multinomial data vector \(y_{ij} = (y_{ij1}, y_{ij2}, y_{ij3}, y_{ij4}, y_{ij5})\), where \(y_{ij} \sim MN(1, p_{ij})\) (Hooten et al., 2010a). The multinomial vectors were constructed using the function \(y_{ij} = h(\{\mu_j(t), \forall t\})\) based on the imputed path realizations by coding a transition as either a stay or a move in a certain direction according to the schematic in Figure B1.

![Figure B1](image-url)

**Figure B1.** Discrete set of possible transitions at any time \(t\), used to create the multinomial vector \(y(t)\), based on the function \(h(\{\mu_j(t), \forall t\})\). (a) Move up: \(y(t) = (1, 0, 0, 0, 0)\). (b) move right: \(y(t) = (0, 1, 0, 0, 0)\). (c) stay: \(y(t) = (0, 0, 1, 0, 0)\). (d) move down: \(y(t) = (0, 0, 0, 1, 0)\). and (e) move left: \(y(t) = (0, 0, 0, 0, 1)\).

Hanks et al. (2015a) reparameterized the multinomial imputation data using sufficient statistics. They denoted residence time as \(\tau_{lj}\) (approximated by \(\Delta t\) times the number of consecutive stays in the current grid cell, Figure A1(c)) for \(l = 1, \ldots, L\) “stay–move” pairs and then defined the probability of staying in the current grid cell for time \(\tau_{lj}\) as \(p_{\tau_{lj}}/\Delta t = (1 - p_{lj, move})\tau_{lj}/\Delta t\), where \(p_{lj, move}\) is the probability of moving. Hanks et al. (2015a) let \(p_{lj, move} = \Delta t \cdot \lambda_{lj, move}\) and \(\Delta t \to 0\) yielding

\[
\lim_{\Delta t \to 0} (1 - p_{lj, move})\tau_{lj}/\Delta t = e^{-\tau_{lj} \lambda_{lj, move}}
\]

which implies that \(\tau_{lj} \sim \text{Exp}(\lambda_{lj, move})\).

Similarly, Hanks et al. (2015a) showed that the movement probability to neighboring grid cell \(c\) is \(p_{clj}/p_{lj, move} = \lambda_{clj}/\lambda_{lj, move}\). Thus, combining the residence probability model with the movement probability yields a likelihood for the sufficient statistic \((\tau_{lj}, y_{lj1}, y_{lj2}, y_{lj4}, y_{lj5})\) equal to \(\prod_{l=1}^{L} \prod_{c \neq j} \lambda_{clj} \text{exp}(-\tau_{lj} \lambda_{clj})\). The likelihood for this reparameterized CTDS model coincides with a Poisson where \(\lambda_{clj}\) is the movement rate to neighboring cell \(c\) and \(\tau_{lj}\) is an offset. Thus, any software capable of fitting a Poisson generalized linear model with an offset can fit the CTDS model if the true path is observed at a fine enough temporal resolution.

Hanks et al. (2015a) used a multiple imputation approach to account for the uncertainty in the path distribution based on (28). The movement rates can then be linked to the environmental covariates by a log-linear link \(\lambda_{clj} = X_{clj}' \beta_j\), where the covariates \(X_{clj}\) can be specified in several meaningful ways to capture either differential movement rates (i.e., motility) or gradient-based directional bias in movement relative to environmental covariates (see Hanks et al., 2015a for details). The reparameterized CTDS model of Hanks et al. (2015a) is much more computationally efficient than that of Hooten et al. (2010a) because the dimensionality of the data \(4L\) depends on the grid cell size instead of the temporal discretization of the path.